

VU Research Portal

High-throughout transcriptome sequencing: data analysis and methods development of large expression data sets

Tang, D.T.P.

2015

document version

Publisher's PDF, also known as Version of record

[Link to publication in VU Research Portal](#)

citation for published version (APA)

Tang, D. T. P. (2015). *High-throughout transcriptome sequencing: data analysis and methods development of large expression data sets*. [PhD-Thesis – Research external, graduation internal, Vrije Universiteit Amsterdam].

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

E-mail address:

vuresearchportal.ub@vu.nl

Nederlandse Samenvatting

Recente ontwikkelingen op het gebied van DNA sequentie analyse hebben het mogelijk gemaakt de DNA code van in principe alle organismen op te helderen. Nieuwe technologieën, waaronder de zogenaamde “high-throughput sequencing”, heeft DNA sequentie analyse mogelijk gemaakt die zowel tijds- als kostenbesparend werkt en deze ontwikkelingen hebben sequencing van volledige genomen, inclusief dat van de mens voor onderzoek mogelijk gemaakt.

Met het gereedkomen van de sequentie van het humane genoom en de sequentie van de genomen van andere species, is de nadruk bij genoom sequentie analyse komen te liggen op de identificatie van kleine functionele elementen in het DNA. Het veld van transcriptoom analyse heeft zich voornamelijk gericht op de correcte annotatie en data analyse van de getranscribeerde producten. De introductie van high-throughput sequencing in transcriptoom analyse heeft een niet vooringenomen en accurate screening van het transcriptoom mogelijk gemaakt, en heeft daarmee inzicht geboden in de grote complexiteit van bijvoorbeeld het humane genoom. Echter, alle methoden zijn nog relatief nieuw en niet gematureerd, eigenlijk allemaal ontwikkeld in de laatste 5-6 jaar. Daarom vinden er continu verbeteringen aan deze methoden plaats. In mijn dissertatie heb ik daarom veel aandacht besteed aan het gebruik van bio-informatica methoden om het transcriptoom beter te kunnen analyseren.

In hoofdstuk 2 richt ik mij daarom op de mogelijke bias en technische artefacten die bij de transcriptoom analyse voor kunnen komen, in het bijzonder die bij de “nano-CAGE” analyse kunnen optreden. In het bijzonder vond ik artefacten die werden geïntroduceerd door het gebruik van moleculaire barcodes en ik heb verschillende methoden ontwikkeld om hier om op een juiste wijze mee om te gaan.

In hoofdstuk 3 heb ik mij gericht op het analyseren van het transcriptoom van cellen met geïnduceerde DNA schade gebruik makend van sequentie analyse van “small RNAs”. Hierbij ontdekten we een nieuwe vorm van RNA molekulen, die worden gevormd dicht op de plaatsen van DNA schade. Deze RNAs vormen de respons op de DNA schade.

In hoofdstuk 4 heb ik de expressie van een specifieke klasse van “small RNAs” bestudeerd, de zogenaamde Piwi-interacterende RNAs (piRNAs). Dit werd gedaan in muizen met een deletie van het Mecp2 gen. Onze analyse liet zien dat piRNAs tot overexpressie komen in het cerebellum van muizen met een Mecp2 deletie, hetgeen te maken kan hebben met de voorgestelde rol van Mecp2 in het silenceren van transposons.

In hoofdstuk 5 heb ik me gericht op het bestuderen van transcripten die ontstaan uit repetitieve elementen (RE) in het DNA. Ik heb daarbij het expressie profiel van deze transcripten voor een groot aantal cellijnen, weefsels, en primaire cellen in kaart gebracht. Ik heb daaruit geconcludeerd dat RE's belangrijk zijn voor de expressie van de “long non-coding RNAs” en “enhancer RNAs”, waarbij hun expressie weefsel specifiek is.

In conclusie, high-throughput sequentie analyse heeft het veld van transcriptoom analyse wezenlijk veranderd door veel meer complexiteit van transcriptie aan te tonen dan daarvoor voor mogelijk werd gehouden.

De interpretatie van deze complexiteit behoort tot een controversieel onderwerp. Aan één einde van het spectrum bestaat de theorie dat het overgrote deel van het genoom bestaat uit functionele onderdelen, terwijl aan de andere kant de mening bestaat dat het genoom voor het merendeel niet functionele producten genereert. Een belangrijk aspect van deze discussie is dat duidelijk is geworden dat een expressie gebeurtenis (transcript) alleen niet genoeg bewijs is voor functionaliteit.

De lagen van complexiteit die we daarbij hebben ontdekt omvatten meerdere klassen van RNA's, ieder met een eigen potentiële regulatoire rol, bijvoorbeeld, random transcriptionele gebeurtenissen ook wel "random noise" genoemd, en de technische artefacten. Het scheiden van technische variatie van het echte signaal vergt begrip van de verschillende technologieën, diepgaande data analyse, en de ontwikkeling van adequate bio-informatica methoden.